

Lawrence Berkeley National Laboratory

Recent Work

Title

Renormalization group computation of likelihood functions for cosmological data sets

Permalink

<https://escholarship.org/uc/item/1d23w885>

Journal

Physical Review D, 99(4)

ISSN

2470-0010

Author

McDonald, P

Publication Date

2019-02-15

DOI

10.1103/PhysRevD.99.043538

Peer reviewed

Renormalization group computation of likelihood functions for cosmological data sets

Patrick McDonald^{1,*}

¹*Lawrence Berkeley National Laboratory, One Cyclotron Road, Berkeley, CA 94720, USA*

(Dated: October 22, 2018)

I show how a renormalization group (RG) method can be used to incrementally integrate the information in cosmological large-scale structure data sets (including CMB, galaxy redshift surveys, etc.). I show numerical tests for Gaussian fields, where the method allows arbitrarily close to exact computation of the likelihood function in order $\sim N$ time, even for problems with no symmetry, compared to N^3 for brute force linear algebra (where N is the number of data points – to be fair, methods already exist to solve the Gaussian problem in at worst $N \log N$ time, and this method will not necessarily be faster in practice). The method requires no sampling or other Monte Carlo (random) element. Non-linearity/non-Gaussianity can be accounted for to the extent that terms generated by integrating out small scale modes can be projected onto a sufficient basis, e.g., at least in the sufficiently perturbative regime. The formulas to evaluate are straightforward and require no understanding of quantum field theory, but this paper may also serve as a pedagogical introduction to Wilsonian RG for astronomers.

I. INTRODUCTION

Structure in the Universe starts as a near perfect Gaussian random field, and remains tightly connected to the initial conditions on large scales [1], motivating using a rigorous likelihood analysis to extract information about models from observed large scale fluctuations. Unfortunately, the likelihood function one can write [e.g., 2] is generally cumbersome to evaluate numerically for realistic large datasets, with time to compute naively scaling like N^3 using brute force linear algebra (where N is the number of data points), which quickly becomes hopeless. Various methods have been devised to speed this up to at worst $N \log N$ [e.g., 3–9], but in practice well-motivated but not explicitly likelihood-based power/correlation estimation methods are usually used instead [e.g., 10, 11]. Here I describe what seems like a particularly elegant way to make likelihood computation fast enough to be practical, based on renormalization group (RG) ideas from quantum field theory [12, 13].

To be clear, the purpose of this paper is not to argue that this is the best/fastest way to estimate power spectra of Gaussian fields. For example, many will inevitably consider the approach of defining semi-heuristic “summary statistics,” like any reasonably computed power spectrum or correlation function of the data, backed up by calibration and error estimation using mock data, to be more appealing to code, surely faster, and optimal enough for their purposes. No doubt this is true a lot of the time. The point here is to present a theoretical physics-oriented framework for viewing the LSS analysis problem, which may present new possibilities for extension to, e.g., the largest scales where pseudo- C_ℓ -type methods are least applicable, and a new way to approach non-linearities. The spirit of this introductory paper is summarized by the old physics saying that “any good theorist can tell you several different ways to solve the same problem.”

While it isn’t at all difficult once you get into it, the math here may look a little bit technical at first glance. The basic idea is straightforward though: First, we effectively integrate over the highest k Fourier modes, i.e., smallest scale structure in the problem, absorbing the numbers that result from this integration into redefinitions of constants in the likelihood function. (For an extreme pedagogical example, suppose we were asked to do the integral $I = A \int dx \int dy f(x)g(y)$, and we knew how to do the integral over y to produce a number $I_y \equiv \int dy g(y)$. Obviously we can do the integral over y , define $A' \equiv AI_y$, and now write the original integral as $I = A' \int dx f(x)$. Trivial as it seems, this is basically renormalization in a nutshell.) Operationally, once the smallest scale structure is gone (already accounted for), we can “coarse grain”, i.e., combine small cells in the data set into bigger cells. This coarse graining is critical because we can’t integrate over all Fourier modes at once – that takes us back to the impossible problem we are trying to get around. After coarse graining, we can integrate out the smallest remaining structure again, then coarse grain further, and so on. It may be puzzling that you will not see any explicit integration over modes in the paper, even though that is presented as the key idea in this kind of RG procedure. Explicit integration over modes is generally replaced by suppression of their fluctuations for technical reasons [12, 13], but with enough thought one can see that these things are effectively equivalent.

* PVMcDonald@lbl.gov

II. EFFECTIVELY EXACT GAUSSIAN LIKELIHOOD EVALUATION

We will primarily discuss Gaussian likelihood functions, where it is easy to make the calculations effectively exact, although the basic RG equation we derive will be exact in principle even for non-Gaussian likelihood functions (with non-Gaussianity, approximations are usually needed to evaluate it).

Suppose we have an observed data vector \mathbf{o} , linearly related to an underlying Gaussian field of interest, ϕ , by $\mathbf{o} = \mu + \mathbf{R}\phi + \epsilon$, where μ is the mean vector, \mathbf{R} is a short-range response matrix and ϵ is Gaussian error with covariance matrix \mathbf{N} , assumed to contain at most short-range correlation. Assume \mathbf{o} has N_o elements and ϕ has N_ϕ , i.e., \mathbf{R} is an $N_o \times N_\phi$ matrix, where fundamentally we should take $N_\phi \rightarrow \infty$ because ϕ is a continuous field (in practice in this paper I leave $N_\phi = N_o$ to start the computation – from there a key point of the RG method is to reduce this by coarsenings). For now assume μ is known and subtract it from \mathbf{o} (the mean or its parameters can be fit for using the method we discuss). The covariance matrix (power spectrum in Fourier basis) of ϕ is $\mathbf{P}(\theta)$, which depends on parameters θ which we are ultimately trying to measure. (In the calculations here I will assume the signal field can be treated as translation invariant so \mathbf{P} is truly diagonal in Fourier space, however, this is not strictly true for the large-scale evolving Universe. After working through the idealized case, it will become clear how to generalize to a less idealized real Universe.) Dropping irrelevant Bayesian factors relating $L(\mathbf{o}|\phi)L(\phi|\theta)$ to $L(\phi, \theta|\mathbf{o})$, the likelihood function for θ and ϕ given \mathbf{o} is

$$L(\phi, \theta|\mathbf{o}) = \frac{e^{-\frac{1}{2}\phi^t \mathbf{P}^{-1} \phi - \frac{1}{2}(\mathbf{o} - \mathbf{R}\phi)^t \mathbf{N}^{-1}(\mathbf{o} - \mathbf{R}\phi)}}{\sqrt{\det(2\pi\mathbf{P}) \det(2\pi\mathbf{N})}}. \quad (1)$$

We are really only interested in $L(\theta|\mathbf{o})$, so the standard approach is to perform the Gaussian integral over ϕ to obtain:

$$L(\theta|\mathbf{o}) = \int d\phi L(\phi, \theta|\mathbf{o}) = \frac{e^{-\frac{1}{2}\mathbf{o}^t \mathbf{C}^{-1} \mathbf{o}}}{\sqrt{\det(2\pi\mathbf{C})}} \quad (2)$$

where $\mathbf{C} = \mathbf{R}\mathbf{P}\mathbf{R}^t + \mathbf{N}$. From here we can, e.g., take θ derivatives to define the standard band power estimator [e.g., 2]. For large data sets, however, the problem of numerically evaluating the resulting equations is usually considered too difficult to bother doing (to make a long story short).

The key to the difficulty in the standard approach is that \mathbf{P} is diagonal in Fourier space but dense in real space, while \mathbf{N} generally vice versa. This makes \mathbf{C} dense in both. Computing the inverse or determinant of \mathbf{C} scales like N^3 , making it impossible to do using brute force linear algebra routines for large data sets. The approach of this paper systematically solves this problem. (Alternative solutions include conjugate gradient $\mathbf{C}^{-1}\mathbf{o}$ multiplication combined with a stochastic determinant calculation, and the approach of [8].)

A. General RG equation

Consider the integral

$$I \equiv \int d\phi e^{-S(\phi)} \equiv \int d\phi e^{-\frac{1}{2}\phi^t \mathbf{Q}^{-1} \phi - S_I(\phi) - \mathcal{N}}. \quad (3)$$

If we take $\mathbf{Q}^{-1} = \mathbf{P}^{-1}$, $S_I(\phi) = \frac{1}{2}\phi^t \mathbf{R}^t \mathbf{N}^{-1} \mathbf{R} \phi - \mathbf{o}^t \mathbf{N}^{-1} \mathbf{R} \phi$, and $\mathcal{N} = \frac{1}{2} \text{Tr} \ln(2\pi\mathbf{P}) + \frac{1}{2} \text{Tr} \ln(2\pi\mathbf{N}) + \frac{1}{2} \mathbf{o}^t \mathbf{N}^{-1} \mathbf{o}$, this integral is exactly $L(\theta|\mathbf{o})$. The $\ln \mathbf{P}$ term represents the normalizing factor $\det \mathbf{P}$, which we need to keep around more carefully than usual because it depends on θ . The \mathbf{N} parts of \mathcal{N} could usually be dropped, assuming \mathbf{N} does not depend on parameters. If we generalize to $\mathbf{Q}^{-1} = \mathbf{P}^{-1} \mathbf{W}^{-1} + \mathbf{K}$, then I is the $L(\theta|\mathbf{o})$ that we want if $\mathbf{W} \rightarrow 1$ and $\mathbf{K} \rightarrow 0$. On the other hand, if $\mathbf{K} \rightarrow \infty$, this term completely dominates the integral making it trivial to evaluate, as long as \mathbf{K} is chosen to be diagonal in Fourier basis. The integration could alternatively be cut off by $\mathbf{W} \rightarrow 0$, or we could use \mathbf{K} together with \mathbf{W} to modify the detailed behavior of the cutoff, but in this paper we will mostly assume $\mathbf{W} = 1$ (because keeping the cutoff term distinct from \mathbf{P} initially seemed simpler). Our RG strategy will be to dial \mathbf{K} from 0 to infinity, while redefining constants in the likelihood function to guarantee that the result of the integral does not change. We will parameterize this dialing by a length scale λ , with $\mathbf{K}(k\lambda \ll 1) \rightarrow 0$ and $\mathbf{K}(k\lambda \gg 1) \rightarrow \infty$, i.e., for any given λ smaller scale modes will already be suppressed (i.e., driven to zero variance) by $\mathbf{K} \rightarrow \infty$, while larger scale modes will be untouched with $\mathbf{K} \rightarrow 0$. λ will start at 0, i.e., $\mathbf{K} = 0$ for all modes, and go toward ∞ , i.e., $\mathbf{K} = \infty$ for all modes. Generally we can choose the details of $\mathbf{K}(k, \lambda)$ in any way that looks useful. For the Gaussian problem, the constants we will redefine to preserve the integral value as \mathbf{Q} changes are the quadratic coefficient of

S_I , which we will call $\mathbf{A}(\lambda)$, with $\mathbf{A}(\lambda = 0) = \mathbf{R}^t \mathbf{N}^{-1} \mathbf{R}$, the linear coefficient of S_I , which we will call $\mathbf{b}(\lambda)$, with $\mathbf{b}(\lambda = 0) = \mathbf{R}^t \mathbf{N}^{-1} \mathbf{o}$, and the constant $\mathcal{N}(\lambda)$, with $\mathcal{N}(\lambda = 0) = \frac{1}{2} \text{Tr} \ln(2\pi \mathbf{P}) + \frac{1}{2} \text{Tr} \ln(2\pi \mathbf{N}) + \frac{1}{2} \mathbf{o}^t \mathbf{N}^{-1} \mathbf{o}$. In other words, for the Gaussian problem the renormalized S_I will be $S_I(\lambda) = \frac{1}{2} \phi^t \mathbf{A}(\lambda) \phi - \mathbf{b}^t(\lambda) \phi$. Note, however, that for the rest of this subsection the calculation will apply for general $S_I(\phi)$, i.e., not just quadratic in ϕ .

To preserve the value of the integral as λ changes, we require the derivative with respect to λ to be zero, i.e.,

$$\frac{\partial I}{\partial \ln \lambda} = - \int d\phi \left(\frac{1}{2} \phi^t \frac{\partial \mathbf{Q}^{-1}}{\partial \ln \lambda} \phi + \frac{\partial S_I}{\partial \ln \lambda} + \frac{\partial \mathcal{N}}{\partial \ln \lambda} \right) e^{-S(\phi)} = 0. \quad (4)$$

The most obvious way to change S_I and \mathcal{N} to compensate the change in \mathbf{Q} would be to adjust the quadratic term in S_I (\mathbf{A} in the definition above) to leave the net quadratic term in the integral unchanged, but this would literally accomplish nothing. To find the alternative changes that will effect the integration over ϕ we follow most closely [13]. Note that

$$\frac{1}{2} \phi^t \frac{\partial \mathbf{Q}^{-1}}{\partial \ln \lambda} \phi e^{-\frac{1}{2} \phi^t \mathbf{Q}^{-1} \phi} = -\frac{1}{2} \phi^t \mathbf{Q}^{-1} \mathbf{Q}' \mathbf{Q}^{-1} \phi e^{-\frac{1}{2} \phi^t \mathbf{Q}^{-1} \phi} = -\frac{1}{2} [\partial_\phi^t \mathbf{Q}' \partial_\phi + \text{Tr}(\mathbf{Q}^{-1} \mathbf{Q}')] e^{-\frac{1}{2} \phi^t \mathbf{Q}^{-1} \phi} \quad (5)$$

where we have defined ' to mean $\partial_{\ln \lambda}$, so we have

$$\begin{aligned} \int d\phi \frac{1}{2} \phi^t \frac{\partial \mathbf{Q}^{-1}}{\partial \ln \lambda} \phi e^{-\frac{1}{2} \phi^t \mathbf{Q}^{-1} \phi - S_I(\phi) - \mathcal{N}} &= -\frac{1}{2} \int d\phi e^{-S_I(\phi) - \mathcal{N}} [\partial_\phi^t \mathbf{Q}' \partial_\phi + \text{Tr}(\mathbf{Q}^{-1} \mathbf{Q}')] e^{-\frac{1}{2} \phi^t \mathbf{Q}^{-1} \phi} \\ &= -\frac{1}{2} \int d\phi e^{-\frac{1}{2} \phi^t \mathbf{Q}^{-1} \phi - \mathcal{N}} [\partial_\phi^t \mathbf{Q}' \partial_\phi + \text{Tr}(\mathbf{Q}^{-1} \mathbf{Q}')] e^{-S_I(\phi)} \\ &= \frac{1}{2} \int d\phi \left(\text{Tr} \left[\mathbf{Q}' \left(\frac{\partial^2 S_I}{\partial \phi \partial \phi^t} - \mathbf{Q}^{-1} \right) \right] - \frac{\partial S_I}{\partial \phi^t} \mathbf{Q}' \frac{\partial S_I}{\partial \phi} \right) e^{-S(\phi)} \end{aligned} \quad (6)$$

where we integrated by parts to make the ϕ derivatives act on $e^{-S_I(\phi)}$. To satisfy Eq. (4) we need:

$$S'_I + \mathcal{N}' = \frac{1}{2} \frac{\partial S_I}{\partial \phi^t} \mathbf{Q}' \frac{\partial S_I}{\partial \phi} + \frac{1}{2} \text{Tr} \left[\mathbf{Q}' \left(\mathbf{Q}^{-1} - \frac{\partial^2 S_I}{\partial \phi \partial \phi^t} \right) \right] \quad (7)$$

If we dropped the ϕ -independent pieces, which are normalization factors that are irrelevant if we are only going to compute expectation values with e^{-S} weight, this is a completely standard RG equation [e.g., 13]. For now we do not want to give up on computing exact likelihoods as a function of parameters, so we keep the normalization factors. Note that the derivation is completely general in the sense that we made no assumption about the details of S_I and \mathbf{Q} (e.g., S_I can be an arbitrary, not necessarily quadratic function of ϕ).

B. RG equations for the Gaussian problem

The interesting content comes from expanding the derivatives of S_I . As mentioned above, we define for the Gaussian problem $S_I = \frac{1}{2} \phi^t \mathbf{A} \phi - \mathbf{b}^t \phi$. Plugging this into Eq. (7) gives

$$\frac{1}{2} \phi^t \mathbf{A}' \phi - \mathbf{b}'^t \phi + \mathcal{N}' = \frac{1}{2} (\mathbf{A} \phi - \mathbf{b})^t \mathbf{Q}' (\mathbf{A} \phi - \mathbf{b}) + \frac{1}{2} \text{Tr} [\mathbf{Q}' (\mathbf{Q}^{-1} - \mathbf{A})] \quad (8)$$

Matching coefficients of ϕ gives

$$\mathbf{A}' = \mathbf{A} \mathbf{Q}' \mathbf{A} \quad (9)$$

$$\mathbf{b}' = \mathbf{A} \mathbf{Q}' \mathbf{b} \quad (10)$$

$$\mathcal{N}' = \frac{1}{2} \mathbf{b}^t \mathbf{Q}' \mathbf{b} + \frac{1}{2} \text{Tr} [\mathbf{Q}^{-1} \mathbf{Q}'] - \frac{1}{2} \text{Tr} [\mathbf{A} \mathbf{Q}'] \quad (11)$$

For Gaussian fields, these are the main result equations in the paper. Note that we are assuming that, if \mathbf{N} is sufficiently short range for this approach to work at all, computing \mathbf{N}^{-1} to set the initial conditions for this evolution will not be a limitation.

To gain understanding, we can unpack what the equation $\mathbf{A}' = \mathbf{A}\mathbf{Q}'\mathbf{A}$ is doing. It is equivalent to $\partial\mathbf{A}^{-1}/\partial\ln\lambda = -\mathbf{Q}'$, i.e., $\mathbf{A}^{-1} = \mathbf{Q}_0 - \mathbf{Q} + \mathbf{A}_0^{-1} = \mathbf{P} + \mathbf{N} - (\mathbf{P}^{-1} + \mathbf{K})^{-1}$ (remember that $\mathbf{A}_0^{-1} = \mathbf{N}$, where we are setting $\mathbf{R} = \mathbf{I}$ in this strictly pedagogical paragraph). I.e., \mathbf{A}^{-1} evolves from \mathbf{N} when $\mathbf{K} \rightarrow 0$ to $\mathbf{C} = \mathbf{P} + \mathbf{N}$ when $\mathbf{K} \rightarrow \infty$, i.e., as the calculation runs we will have in some sense $\mathbf{A} = \mathbf{N}^{-1}$ for large scales but $\mathbf{A} = \mathbf{C}^{-1}$ for small scales. The evolution equation for \mathbf{b} has a simple solution in terms of \mathbf{A} : $\mathbf{b} = \mathbf{A}\mathbf{o}$, i.e., given that we understand \mathbf{A} evolution as interpolation from \mathbf{N}^{-1} to \mathbf{C}^{-1} , \mathbf{b} evolution is interpolation from \mathbf{N}^{-1} data weighting to \mathbf{C}^{-1} weighting. Similarly, $\mathcal{N}(\lambda) = \frac{1}{2}\mathbf{o}^t\mathbf{A}\mathbf{o} + \frac{1}{2}\text{Tr}\ln(2\pi\mathbf{A}^{-1}) + \frac{1}{2}\text{Tr}\ln(2\pi\mathbf{Q})$. Noting that \mathbf{Q} evolves from \mathbf{P} to \mathbf{K}^{-1} , we see that this evolves from the initial normalization piece $\mathcal{N}(\mathbf{K} = 0) = \frac{1}{2}\mathbf{o}^t\mathbf{N}^{-1}\mathbf{o} + \frac{1}{2}\text{Tr}\ln(2\pi\mathbf{N}) + \frac{1}{2}\text{Tr}\ln(2\pi\mathbf{P})$ to the final answer $\mathcal{N}(\mathbf{K} \rightarrow \infty) = \frac{1}{2}\mathbf{o}^t\mathbf{C}^{-1}\mathbf{o} + \frac{1}{2}\text{Tr}\ln(2\pi\mathbf{C}) + \frac{1}{2}\text{Tr}\ln(2\pi\mathbf{K}^{-1})$ (where the \mathbf{K} part cancels against the integration term in Eq. 13). The whole point of the paper is that you get to this final answer without actually evaluating these matrix formulas – \mathcal{N} is just an evolving single number, \mathbf{b} is an evolving vector. The key is keeping \mathbf{A} from spreading too far off-diagonal, because formally you do need to store it as a matrix. This is where the coarsening step we will discuss below becomes necessary – \mathbf{A} obviously does spread off-diagonal, but as it does it becomes unnecessary to store it at high resolution, so we can coarse grain to keep the numerically relevant number of cells off-diagonal limited, even as the physical range increases.

Note that for the Gaussian problem we can always formally integrate Eq. 3 over ϕ to obtain

$$L(\theta|\mathbf{o}) = e^{\frac{1}{2}\mathbf{b}^t(\mathbf{Q}^{-1}+\mathbf{A})^{-1}\mathbf{b}-\mathcal{N}}\sqrt{\det\left(2\pi(\mathbf{Q}^{-1}+\mathbf{A})^{-1}\right)}. \quad (12)$$

All of the components will change with λ , but the overall result will not change. Once $\mathbf{K} \gg \mathbf{P}^{-1}$, and $\mathbf{K} \gg \mathbf{A}$, and $\mathbf{b}^t\mathbf{K}^{-1}\mathbf{b} \rightarrow 0$ (all of which we showed above inevitably do happen as $\lambda \rightarrow \infty$) this goes to

$$L(\theta|\mathbf{o}) \xrightarrow{\lambda \rightarrow \infty} e^{\frac{1}{2}\mathbf{b}^t\mathbf{K}^{-1}\mathbf{b}-\mathcal{N}}\sqrt{\det(2\pi\mathbf{K}^{-1})} \rightarrow e^{-\mathcal{N}}\sqrt{\det(2\pi\mathbf{K}^{-1})}, \quad (13)$$

i.e., the final answer is just $e^{-\mathcal{N}}$, up to a $\det\mathbf{K}$ factor which is canceled by a similar factor accumulated within \mathcal{N} , as mentioned above. We have accomplished the integration over ϕ .

C. Cutoff function

Eqs. (9-11) are no good if we can't evaluate them numerically. The key is to arrange \mathbf{Q}' to do its job of suppressing the smallest scale structure to allow coarse graining, but beyond that have as short a real space range as possible (i.e., not coupling widely separated cells), to keep \mathbf{A} , which starts as $\mathbf{R}^t\mathbf{N}^{-1}\mathbf{R}$, as short range as possible (i.e., effectively sparse, when stored in real space). For the above definition we have $\mathbf{Q} = (\mathbf{P}^{-1}\mathbf{W}^{-1} + \mathbf{K})^{-1}$, so

$$\mathbf{Q}' = \mathbf{Q}(\mathbf{P}^{-1}\mathbf{W}^{-1}\mathbf{W}'\mathbf{W}^{-1} - \mathbf{K}')\mathbf{Q}. \quad (14)$$

If $\mathbf{K} = 0$ this is just $\mathbf{Q}' = \mathbf{W}'\mathbf{P}$. On the other hand, if $\mathbf{W} = 1$ we have $\mathbf{Q}' = -\mathbf{Q}\mathbf{K}'\mathbf{Q}$. Let's focus on the $\mathbf{W} = 1$ case, and assume \mathbf{K} will not depend on the power spectrum as a function of parameters of interest (it might depend on a fiducial power spectrum that would be kept fixed when varying parameters). Making \mathbf{K} a simple function of k in Fourier space, we can efficiently compute \mathbf{Q}' in real space, which takes the form of a convolution kernel (meaning we don't need to compute it as a full matrix, not even a sparse one – we can compute it as a function of separation and use translation invariance to move it around).

We are left to choose a specific \mathbf{K} . Suppose \mathbf{P} is a constant P in Fourier space and we choose $K(k, \lambda) = P^{-1}k^2\lambda^2\exp(k^2\lambda^2)$ which accomplishes the goal of dominating (going to ∞) at $k\lambda \gg 1$ and disappearing (going to zero) at $k\lambda \ll 1$, then $Q'(x = k\lambda) = -2P\frac{x^2(1+x^2)\exp(x^2)}{(1+x^2\exp(x^2))^2}$. This is a fairly nice function that goes like x^2 at small x , peaks at $x \sim 1$, and goes like $\exp(-x^2)$ at large x – it should produce a nice limited range convolution kernel. Another very simple example is $K(k, \lambda) = P^{-1}(k)\lambda^2k^2$. This gives \mathbf{Q}' going like $P(k)k^2$ at $k\lambda \ll 1$ and $P(k)k^{-2}$ at $k\lambda \gg 1$. We see that the function we choose is picking the detailed range of $P(k)$ to emphasize for a given λ (when varying parameters, the $P(k)$ in K would stay fixed). In the numerical example below we will use $K(k, \lambda) = P_0^{-1}(k)(k\lambda)^\alpha\exp[(k\lambda)^2]$, with $\alpha = 2$ or 2.5 , where $P_0(k)$ is the fiducial power spectrum. Tuning the cutoff function is a topic for more thought and experimentation.

D. Coarse graining

In any case, as λ increases successive convolutions with \mathbf{Q}' in general extend the correlation range of \mathbf{A} , eventually making it the dense matrix (\mathbf{C}^{-1} actually, as mentioned above) that we want to avoid, if we don't do anything about

TABLE I. Evolution of various quantities from start to asymptotic endpoint (for $\mathbf{R} = \mathbf{I}$). \mathbf{P} and \mathbf{K} , and thus \mathbf{Q} and \mathbf{Q}' , are diagonal in Fourier space, while \mathbf{N} is at least approximately diagonal in real space, and \mathbf{A} remains so after coarse graining (it would become dense without coarse graining because $\mathbf{C} \equiv \mathbf{P} + \mathbf{N}$ is dense).

quantity	start ($k\lambda \ll 1$)	\rightarrow	end ($k\lambda \gg 1$)
λ	0		∞
\mathbf{K}	0		∞
\mathbf{Q}	\mathbf{P}		$\mathbf{K}^{-1} \rightarrow 0$
\mathbf{Q}'	0		0 ^a
\mathbf{A}	\mathbf{N}^{-1}		\mathbf{C}^{-1} b
\mathbf{b}	$\mathbf{N}^{-1}\mathbf{o}$		$\mathbf{C}^{-1}\mathbf{o}$ b
\mathcal{N}	$\frac{1}{2}\mathbf{o}^t\mathbf{N}^{-1}\mathbf{o} + \frac{1}{2}\text{Tr}\ln(2\pi\mathbf{N}) + \frac{1}{2}\text{Tr}\ln(2\pi\mathbf{P})$		$\frac{1}{2}\mathbf{o}^t\mathbf{C}^{-1}\mathbf{o} + \frac{1}{2}\text{Tr}\ln(2\pi\mathbf{C})$ c

^a non-zero for $k\lambda \sim 1$, see Figs. 1 and 2

^b coarse grained

^c det \mathbf{K} term canceled

it (to be clear: evaluating Eq. 9 scales like the N^3 we said was too large if \mathbf{A} becomes dense). At the same time, however, the growth of \mathbf{K} suppressing high- k fluctuations in ϕ makes resolution much finer than approximately the same scale unnecessary, i.e., we should be able to coarsen the pixelization to reduce the range of \mathbf{A} as measured in cells off-diagonal. This can be seen by imagining Eq. (12) in Fourier basis. At high k where $\mathbf{K} \rightarrow \infty$, any other structure is overwhelmed so the full simplification of Eq. (13) applies, i.e, apparently the values of \mathbf{b} and \mathbf{A} in these high- k modes are irrelevant. All of the integral contribution in high- k modes must now be contained in \mathcal{N} – this is where we see that the RG evolution equation is equivalent to integrating over these modes. Since we know that high k modes in \mathbf{b} and \mathbf{A} do not contribute to the remaining integral, we can compress them to make the computations faster. Another way to see this is: all of the terms in the evolution equations (Eqs. 9-11) involve multiplication of \mathbf{A} or \mathbf{b} by \mathbf{Q}' , but by construction \mathbf{Q}' has no structure on scales much less than λ (i.e., is zero in Fourier space). This means that, operationally, it is clear that \mathbf{A} and \mathbf{b} can be compressed at some point, i.e., their small scale structure is irrelevant when it is multiplied by a smooth function. Imagining 1D, the compression process can be simply $b_1^c = b_1 + b_2$, $b_2^c = b_3 + b_4$, $A_{11}^c = A_{11} + 2A_{12} + A_{22}$, etc., where \mathbf{b}^c and \mathbf{A}^c are the new objects compressed by a factor of 2 in each index. Understanding that on small scales $\mathbf{b} \rightarrow \mathbf{C}^{-1}\mathbf{o}$ and $\mathbf{A} \rightarrow \mathbf{C}^{-1}$, we see that this process is intuitively straightforward, equivalent to an optimally weighted averaging of adjacent cells.

Especially in higher dimensions, the problem is greatly simplified after even a single factor of 2 in coarsening. Eventually the data set contains few enough cells to just evaluate the remaining integration by brute force (i.e., plug \mathbf{A} , etc. that we have at that point into Eq. 12 and compute using standard linear algebra). To parallelize the algorithm, we could split the data set by processor, where for non-shared memory, each processor would need enough extra data around its local patch to account for the convolution from outside.

For a summary of the properties of various quantities we have defined, see Table I.

E. Numerical example

Eqs. (9-11) are exact for the Gaussian problem, so in some sense they must work, but it is helpful to work through a numerical example to reassure ourselves that they are practical. For coding simplicity, I work in one dimension, in units of the basic data cell size. I include unit variance noise in each cell. I use signal power spectrum $P(k) = A(k/k_p)^\gamma \exp(-k^2)$ with $\gamma = 0$ or -0.5 . I generate mock data with $A_0 = 1$ and calculate likelihoods as a function of A . I use pivot $k_p = 0.1$ to make the constraining power on A roughly similar between the two γ s (for my largest datasets). I choose the cutoff function $K(k, \lambda) = P_0^{-1}(k)(k\lambda)^\alpha \exp[(k\lambda)^2]$, where $P_0(k)$ is the fiducial power spectrum. I initially used $\alpha = 2$, but found that $\alpha = 2.5$ seems to be somewhat more efficient for the $\gamma = -0.5$ case. In Figure 1 I show the critical \mathbf{Q}' in Fourier space (for $P(k) = 1$, to isolate the scale dependence of \mathbf{K} without mixing it with $P(k)$). For any λ , this shows the range of k that is being “integrated out.” The real space version of \mathbf{Q}' is shown in Figure 2. It has mean zero, meaning it does not affect scales much larger than λ – this is what keeps \mathbf{A} short range – and it is smooth so that it does not affect scales much smaller than λ – this is why we can coarsen \mathbf{A} and \mathbf{b} .

For small problems, we can evolve Eqs. (9-11) using a dense matrix for \mathbf{A} , and check step by step that the components and final likelihood agree with the analytic solutions. I find that steps $d\ln\lambda = 0.2$ are sufficient in practice for my test problems, after realizing that it is very helpful to use at least 2nd order accurate mid-point stepping (https://en.wikipedia.org/wiki/Midpoint_method), instead of the most naive possible 1st order stepping. For

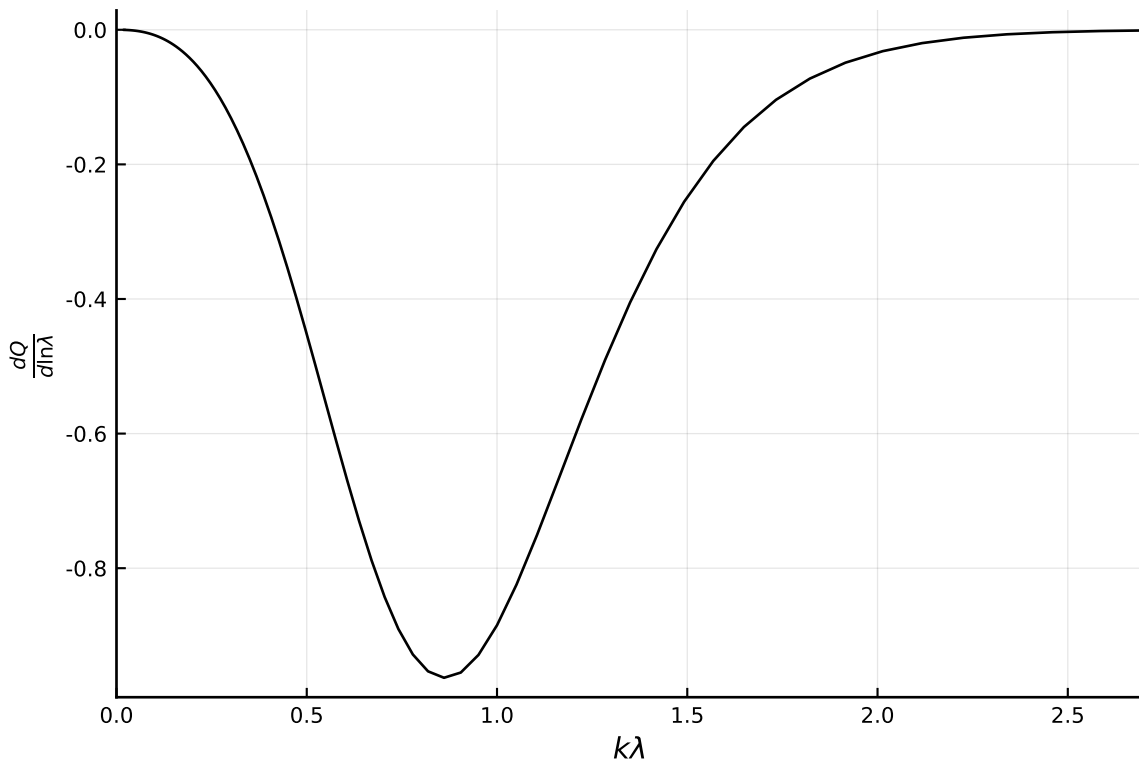


FIG. 1. Fourier space $\frac{dQ}{d \ln \lambda}(k)$ vs. $k\lambda$. This shows the range in which fluctuations in ϕ are being “integrated out”.

larger problems I use canned Julia sparse matrix operations for \mathbf{A} . Since \mathbf{Q}' is formally dense (just falling rapidly to zero with separation), we need to impose by hand a cut on the range of matrix operations to store, which I set after experimentation to 20λ . Note that this, like $d \ln \lambda$ and other numerical parameters of the method, is a highly controlled approximation that can be checked for convergence without reference to any known answer. The final non-trivial numerical parameter is when to coarsen, which I set after experimentation to when λ has evolved to be 7 times the current cell size, i.e., the formula is “if $\lambda > 7c$, $c \rightarrow 2c$ ”, where c is the cell size. These settings determine the overall speed of the analysis, because they determine how many cells off diagonal must be stored in \mathbf{A} , i.e., the product of the two numbers tells us how efficiently we are accomplishing the dual goals of rubbing out small-scale structure enough to coarsen without broadening the range at which we need to store correlations any more than necessary. It may seem like my numbers are surprisingly large, but I am requiring quite high accuracy in the likelihood calculation, and did not try very hard to, e.g., tune the cutoff function to potentially reduce them. Fig. 3 shows the time per likelihood evaluation, scaling like N_o , compared to brute force linear algebra scaling like N_o^3 . Once the data set is coarsened to 2048 cells I simply evaluate the remaining integral all at once using Eq. (12).

At this point plotting the results is somewhat anti-climactic, because the calculation simply works, more or less by construction. Fig. 4 shows the likelihood computed for an $N_o = 262144$ data set for different power amplitude values, compared to an exact FFT likelihood that I can compute for this data set because I made it periodic with homogeneous noise. The RG calculation does not take advantage of this symmetry in any way. There is one imperfection hidden in Fig. 4, which shows not absolute $2 \ln L$ but its deviation from maximum – the maximum is off by 36 in the RG calculation. This is like if you were doing a χ^2 test of goodness of fit, and got χ^2 wrong by 36 for 262144 degrees of freedom – it is insignificant to the goodness of fit calculation, which has an rms error of 724. The important thing is that the offset is practically the same for all parameter values, so does not affect our conclusions about the parameter (I can make χ^2 more accurate by dialing up numerical settings, but there is no point [14]).

While it is pretty clear that the method will deal gracefully with edges and varying noise, just to be sure I did some tests of this with $N_o = 4096$ data sets, for which I can calculate the exact answer by brute force. One test was to multiply the noise in cells in the second half of the data vector by a factor of 10, and the other was to multiply the noise in even-index (i.e., every other) cells by 10. These tests give similar agreement to what we see in Fig. 4. Note that unobserved cells are equivalent to infinite noise, which enter these calculations nicely as $N_{ii}^{-1} = 0$, which can be dropped from sparse storage of \mathbf{A} . Formally infinite “zero padding” around an observed region, i.e., consistently

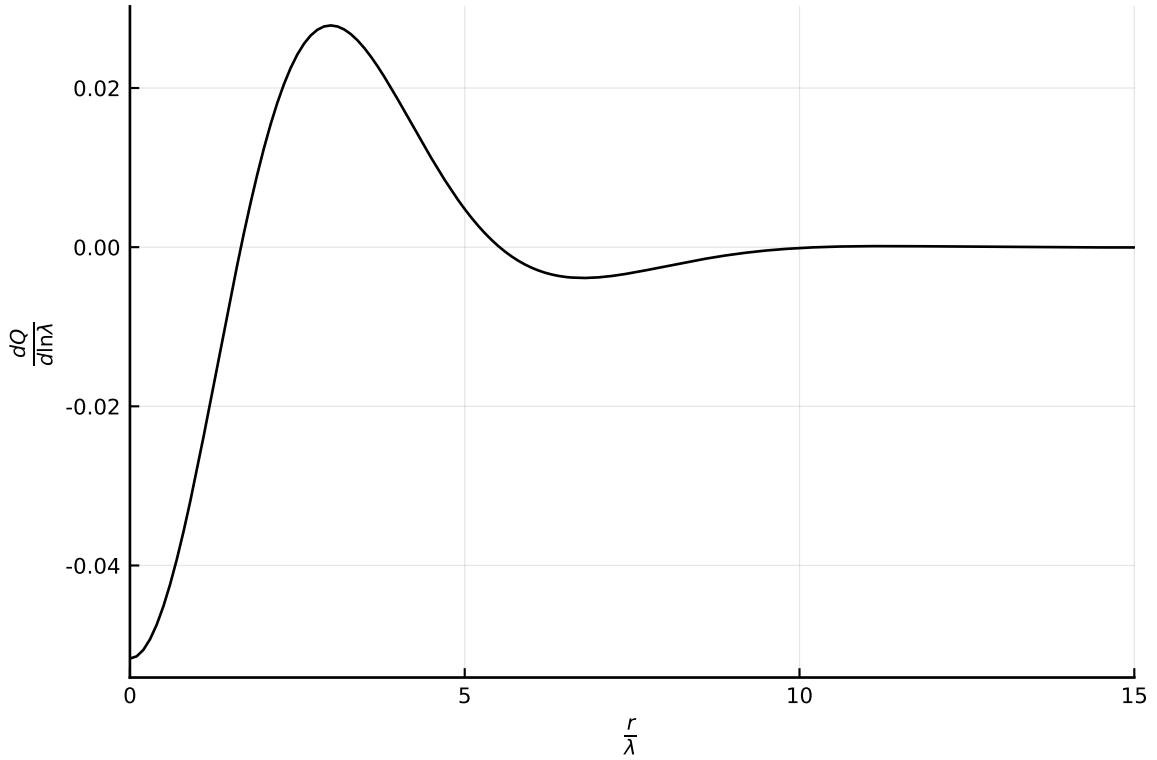


FIG. 2. Real space $\frac{dQ}{d\ln\lambda}(r)$ vs. $\frac{r}{\lambda}$. This shows the convolution kernel applied during the evolution of \mathbf{A} , \mathbf{b} , \mathcal{N} .

implemented unobserved cells, incur no computational cost beyond the range of smearing of \mathbf{A} .

There are some subtleties in numerical evaluation that a potential user should be aware of. You don't want to compute the quantity $(\mathbf{Q}^{-1} + \mathbf{A})^{-1}$ that appears in Eq. 12 naively by computing \mathbf{Q}^{-1} , adding that to \mathbf{A} , and inverting the resulting matrix, because $\mathbf{Q}^{-1} \rightarrow \mathbf{K} \rightarrow \infty$ for high k modes, meaning in practice that, after some evolution, there are elements of \mathbf{Q}^{-1} too big for the computer to handle. The formally equivalent quantity $\mathbf{Q}(\mathbf{I} + \mathbf{A}\mathbf{Q})^{-1}$ is well behaved when used to compute $\mathbf{b}^t (\mathbf{Q}^{-1} + \mathbf{A})^{-1} \mathbf{b}$, because the high k elements of \mathbf{Q} go to zero. This trick does not entirely solve the problem for $\det[(\mathbf{Q}^{-1} + \mathbf{A})^{-1}]$, however because near-zero eigenvalues are a problem for the determinant. At this point we notice, however, that we don't need or want to actually include the term $\text{Tr}[\mathbf{Q}^{-1}\mathbf{Q}']$ in our numerical evolution of \mathcal{N}' (Eq. 11), because we can use the analytic solution of this piece of the equation, $\text{Tr} \ln \mathbf{Q}$, to analytically cancel the problematic $\det \mathbf{Q}$ in $\det[\mathbf{Q}(\mathbf{I} + \mathbf{A}\mathbf{Q})^{-1}]$ in Eq. 12, leaving us to compute the perfectly well-behaved $\det[\mathbf{I} + \mathbf{A}\mathbf{Q}]$.

III. BEYOND-GAUSSIAN LIKELIHOOD

The derivation of Eq. (7) made no assumption about the form of ϕ dependence of S_I – it is valid for general $S_I(\phi)$. In practice S_I can be something other than quadratic due to primordial non-Gaussianity [e.g., 15–20], non-linear gravitational evolution [e.g., 21–44], gravitational lensing [e.g., 45–47], non-linear bias [e.g., 48–62], non-Gaussian noise [e.g., 63], etc. [e.g., 64]. Eq. (7) always applies formally, but the tricky part is that we must be able to match the functional dependence generated on the right hand side against terms on the left hand side. E.g., if we add a ϕ^3 term to $S_I(\phi)$, we will find that all higher order polynomial terms are generated by the evolution. This is due to the $\frac{\partial S_I}{\partial \phi^i} \mathbf{Q}' \frac{\partial S_I}{\partial \phi^j}$ term in Eq. (7), which feeds polynomial terms of order n and m into a term of order $n + m - 2$, i.e., if $n, m \leq 2$, this is never larger than 2, but $n = m = 3$ leads to a fourth order term, and so on. (Fig. 9.2 of [13] nicely represents Eq. 7 in terms of Feynman diagrams.) These higher order terms aren't necessarily relevant though, i.e., computationally significant, so one can hope to simply ignore terms beyond some order. Intuitively, this should be more likely to work if one starts the evolution at relatively low resolution, i.e., on larger, more perturbative scales.

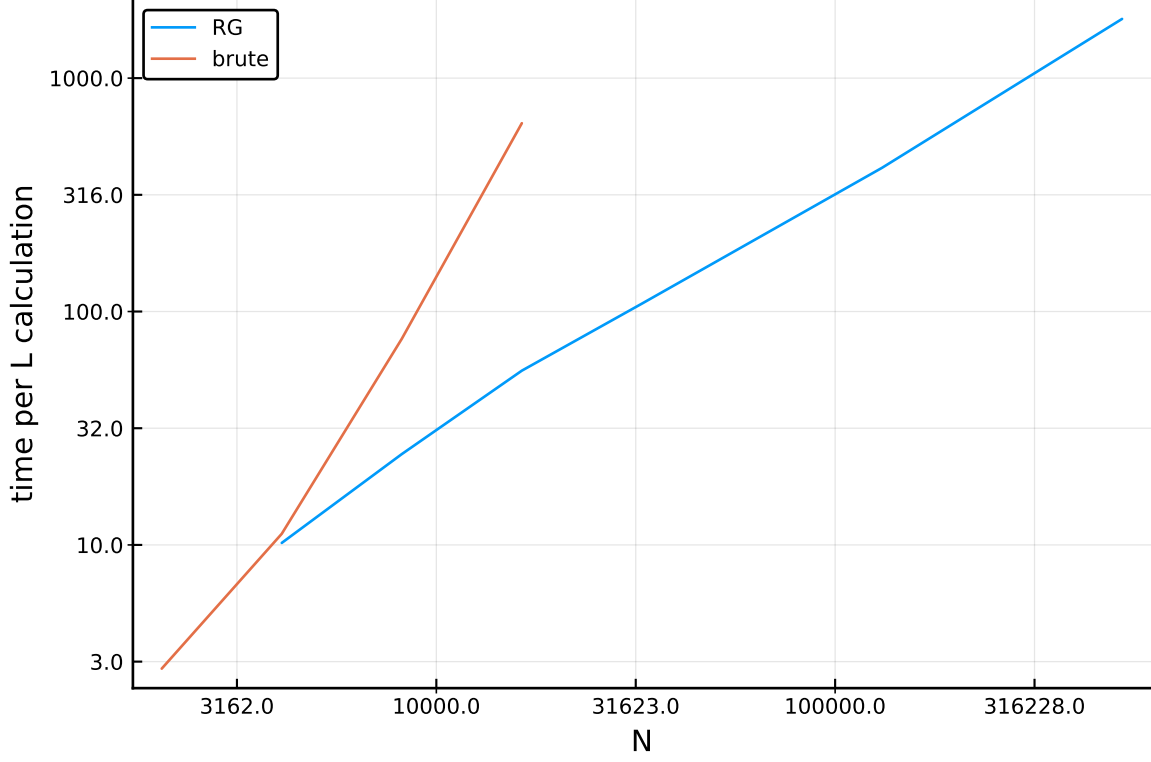


FIG. 3. Time per likelihood evaluation, scaling linearly with N_o for the RG method, compared to N_o^3 for brute force linear algebra. This is implemented using canned Julia sparse matrix routines, without much effort to optimize, so the normalization can probably be reduced significantly. This is running on 1 processor, where, e.g., the $N_o = 524288$ run takes 30 minutes.

Note that a polynomial S_I is not necessarily the best way to represent non-linearity [e.g., 38].

IV. DERIVATIVES OF $\ln L$, I.E., QUADRATIC ESTIMATOR-TYPE APPLICATION

We could always run the above likelihood calculation for different parameter values in order to compute parameter dependence, however, it may be more efficient to derive the analytic formula for derivatives with respect to parameters and then compute that from the data, in the usual way of “quadratic estimators.” Derivatives with respect to parameters bring down quadratic factors in ϕ . We should be able to similarly accumulate these integrals.

Consider

$$\frac{\partial \ln L(\theta|\mathbf{o})}{\partial \theta_\alpha} = -\frac{1}{2} \text{Tr} [\mathbf{P}^{-1} \mathbf{P}_{,\alpha}] + \frac{1}{2} \frac{\int d\phi \phi^t \mathbf{P}^{-1} \mathbf{P}_{,\alpha} \mathbf{P}^{-1} \phi e^{-\frac{1}{2} \phi^t \mathbf{P}^{-1} \phi - S_I(\phi)}}{\int d\phi e^{-\frac{1}{2} \phi^t \mathbf{P}^{-1} \phi - S_I(\phi)}} \quad (15)$$

where $\mathbf{P}_{,\alpha} \equiv \partial \mathbf{P} / \partial \theta_\alpha$. Note that we do not need to keep track of the common normalization factor between the numerator and denominator. We now look to accumulate the contribution coming from insertion of $\frac{1}{2} \phi^t \mathbf{P}^{-1} \mathbf{P}_{,\alpha} \mathbf{P}^{-1} \phi$. The leading trace factor can be computed trivially in Fourier space.

Noting that $f[\phi] e^{-S(\phi)} = e^{-S(\phi) + \ln f[\phi]}$ we see that we can compute integrals like in Eq. (15), i.e., integrals of the form $\int d\phi f[\phi] e^{-\frac{1}{2} \phi^t \mathbf{P}^{-1} \phi - S_I(\phi)}$, by simply inserting $S_I - \ln f$ in place of S_I in Eq. (7). In this way we derive, in addition to the same evolution equations for S_I (i.e., \mathbf{A} and \mathbf{b}) above,

$$f' = \frac{\partial f}{\partial \phi^t} \mathbf{Q}' \frac{\partial S_I}{\partial \phi} - \frac{1}{2} \text{Tr} \left[\frac{\partial^2 f}{\partial \phi \partial \phi^t} \mathbf{Q}' \right]. \quad (16)$$

Note that this is effectively a special case of the general non-Gaussianity discussed in §III, where the particular form of effective non-Gaussianity here guarantees that we have an exact basis to project onto.

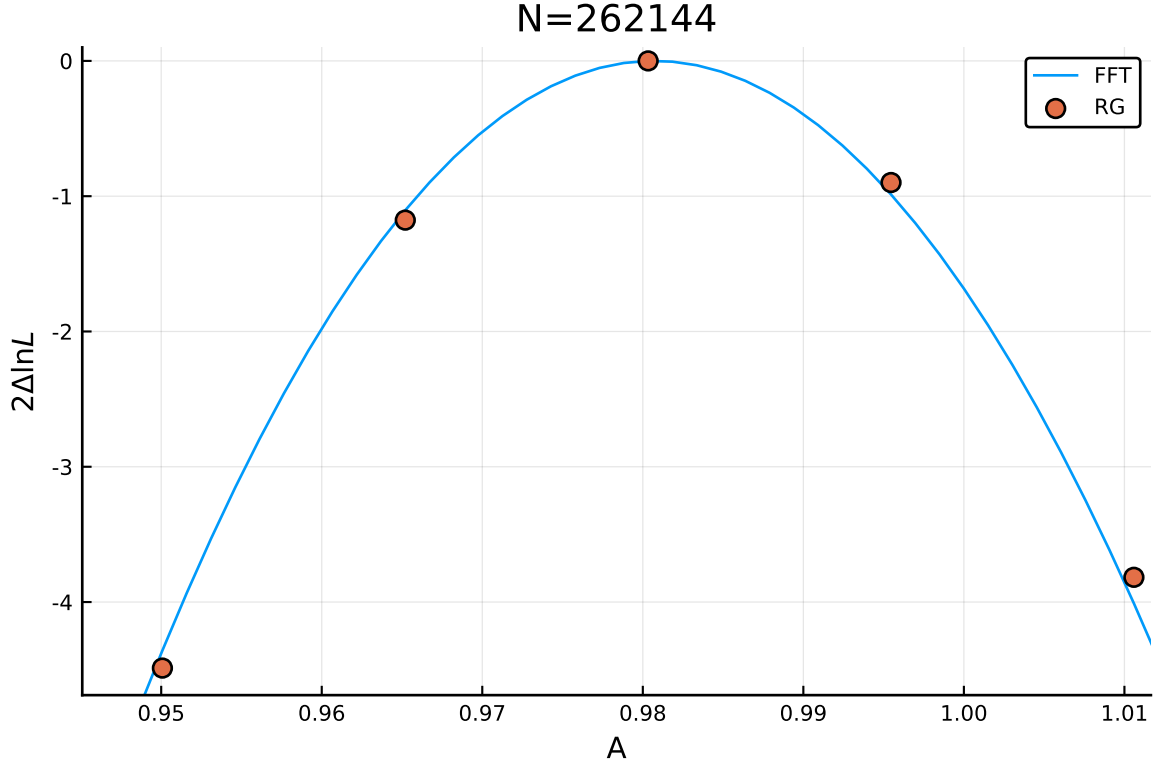


FIG. 4. RG likelihood for five values of the power amplitude, compared to the exact FFT likelihood that I can compute because this data set was periodic with homogeneous noise. This is for $\gamma = -0.5$, $N = 262144$ data points. All similar figures look similar, by construction, because if they didn't I would have dialed up the numerical accuracy parameters until they did.

In the case of Eq. (15), we have $f_\alpha[\phi, \lambda] \equiv \frac{1}{2}\phi^t \mathbf{P}^{-1} \mathbf{P}_{,\alpha} \mathbf{P}^{-1} \phi + \frac{1}{2}\phi^t \mathbf{D}_\alpha(\lambda) \phi + \mathbf{c}_\alpha(\lambda) \phi + E_\alpha(\lambda)$, where $\mathbf{D}_\alpha(\lambda = 0) = 0$, $\mathbf{c}_\alpha(\lambda = 0) = 0$, and $E_\alpha(\lambda = 0) = 0$, giving

$$\frac{1}{2}\phi^t \mathbf{D}'_\alpha \phi + \mathbf{c}'_\alpha \phi + E'_\alpha = (\phi^t \mathbf{P}^{-1} \mathbf{P}_{,\alpha} \mathbf{P}^{-1} + \phi^t \mathbf{D}_\alpha + \mathbf{c}'_\alpha) \mathbf{Q}' (\mathbf{A} \phi - \mathbf{b}) - \frac{1}{2} \text{Tr} [\mathbf{P}^{-1} \mathbf{P}_{,\alpha} \mathbf{P}^{-1} \mathbf{Q}'] - \frac{1}{2} \text{Tr} [\mathbf{D}_\alpha \mathbf{Q}'] \quad (17)$$

Matching coefficients of powers of ϕ we find:

$$\mathbf{D}'_\alpha = \mathbf{P}^{-1} \mathbf{P}_{,\alpha} \mathbf{P}^{-1} \mathbf{Q}' \mathbf{A} + \mathbf{D}_\alpha \mathbf{Q}' \mathbf{A} + \text{transpose} \quad (18)$$

(“transpose” here means transpose of the part I wrote on the right hand side to make \mathbf{D}'_α symmetric)

$$\mathbf{c}'_\alpha = \mathbf{A} \mathbf{Q}' \mathbf{c}_\alpha - \mathbf{P}^{-1} \mathbf{P}_{,\alpha} \mathbf{P}^{-1} \mathbf{Q}' \mathbf{b} - \mathbf{D}_\alpha \mathbf{Q}' \mathbf{b} \quad (19)$$

and finally

$$E'_\alpha = -\mathbf{c}_\alpha^t \mathbf{Q}' \mathbf{b} - \frac{1}{2} \text{Tr} [\mathbf{P}^{-1} \mathbf{P}_{,\alpha} \mathbf{P}^{-1} \mathbf{Q}'] - \frac{1}{2} \text{Tr} [\mathbf{D}_\alpha \mathbf{Q}'] \quad (20)$$

Note that once $\lambda \rightarrow \infty$, all expectation values of ϕ go to zero, so the final result of interest is just E_α (which comes with a normalization factor that is canceled by the denominator). The accumulation of \mathbf{D}_α , \mathbf{c}_α , and E_α will be dominated by the range of λ corresponding to their band scale, so it should not be necessary to compute them for all λ , i.e., for a given λ only a limited range of bands near this scale will need to be computed, so the computation time should not increase by too large a factor over a single likelihood calculation.

The key to the efficiency of this calculation appears to be keeping \mathbf{D}_α short range, as we discussed for \mathbf{A} above. The key to this is the quantity $\mathbf{P}^{-1} \mathbf{P}_{,\alpha} \mathbf{P}^{-1} \mathbf{Q}'$ being a short-range convolution. The \mathbf{Q}' is no more problem than discussed above, but $\mathbf{P}_{,\alpha}$ brings up a new worry. If this represents a band in power narrower than \mathbf{Q}' , the range of convolution must be inversely proportional to the width of the band, i.e., increasing k resolution in the power spectrum measurement will generally require extended range of \mathbf{D}_α . As long as the required fractional resolution is reasonable, this will just be a prefactor on the fast $\mathcal{O}(N)$ scaling.

The 2nd derivative of $\ln L$ will proceed similarly, with two-index objects to accumulate.

V. DISCUSSION

Obvious next steps include applying this approach to real data analysis, and thinking further beyond Gaussianity. Note that the definition of \mathbf{Q} does not need to be isotropic. E.g., for a spectroscopic redshift survey like DESI overlapping with a lensing survey like LSST [65, 66], we may want to first integrate out small-scale radial modes to which the lensing survey is not sensitive, before transitioning to a joint analysis of primarily transverse modes. \mathbf{Q} can and should be tuned to the problem at hand, with the only requirement being that calculations are tractable and have the right limits (i.e., \mathbf{Q} must start as \mathbf{P} and go to zero for any mode that you want to compress away). Note that in greater than one dimension it may be faster to integrate and compress by factors of 2 in one dimension at a time instead of all simultaneously. It might also be possible to compress by less than factors of 2, interpolating onto an intermediate resolution grid instead of simply adding perfectly aligned pairs of cells.

Considering Fig. 2, it seems likely that a more carefully chosen function that had, in particular, only one zero-crossing, could be more efficient. I.e., we need one zero-crossing to have mean zero, but given that we would like the function to be as smooth and compact as possible, and two zero-crossings surely aren't optimal. Now that we understand the method better, we see that it may be simpler to use the multiplicative \mathbf{W} form of cutoff instead of the additive \mathbf{K} . Another possibility might be to give up exact translation invariance in \mathbf{W} or \mathbf{K} and instead choose them to specifically suppress the difference between cells one wants to sum (as opposed to all nearby cells, which is what the translation invariant version does, but may be inefficient).

For the Gaussian problem where we can solve the evolution equations analytically (see §II B), there is an interesting possibility of jumping directly from one coarsening step to the next using $\mathbf{A}^{-1} = \mathbf{A}_0^{-1} + \mathbf{Q}_0 - \mathbf{Q}$, i.e., given \mathbf{A}_0^{-1} after a coarsening, we take the full λ step to the next coarsening by computing $\mathbf{Q}_0 - \mathbf{Q}$ and simply adding – this quantity will have similar sparsity properties to what we have discussed for \mathbf{A} . The trade-off appears to be that now to propagate the contribution to \mathcal{N} we will need to numerically invert and compute the determinant of this \mathbf{A}^{-1} . This should be doable because it is sparse, but may not be any faster than just taking the differential steps as in this paper. Note that I only took 3-4 steps between factors of 2 coarsening in the example problem, with the exception being leading up to the first coarsening (and even there, it is the last steps before the coarsening that take the most time, because \mathbf{A} has spread the most). This idea would probably be most interesting if in some scenario it is found that significantly finer $\ln \lambda$ steps are required.

This type of RG approach has been applied to calculations of data-independent statistics of the evolving density field [e.g., 67–71], but the very simple application here may cast new light on how to think about those applications. [38] provides an exact functional integral starting point for this. This RG approach should also be another tool within the “information field theory” formalism of [63, 72].

A fully dense, arbitrary, inhomogeneous noise matrix will of course confound this method. A translation invariant noise component, i.e., diagonal in Fourier space, could be included in the signal power spectrum part of the calculation. A contaminant that can be modeled by a reasonable number of vectors contributing to the mean of the field can be marginalized over (with simplifications if the vectors are, e.g., smooth or localized).

Finally, asking the question “what will we do with a redshift survey of the real, evolving (i.e., not translation invariant) Universe” leads to the following re-understanding of the key points of the method: Take the definition $\mathbf{Q} \equiv \mathbf{W}\mathbf{S}$, where I use \mathbf{S} here where I had used \mathbf{P} before to emphasize that the signal covariance matrix does not need to be neatly diagonalizable by a Fourier transform. \mathbf{W} starts at \mathbf{I} , while we can remove a mode from the data set if \mathbf{W} drives its variance to zero. This is simplest to visualize and implement if \mathbf{S} and \mathbf{W} can be multiplied along the diagonal in Fourier space, but clearly a \mathbf{W} matrix representing any kind of smoothing operation that suppresses small-scale fluctuations will allow us to coarse grain. The requirement for the calculation to work is simply that we can efficiently compute a compact $\mathbf{Q}' = \mathbf{W}'\mathbf{S}$.

ACKNOWLEDGMENTS

I thank U. Seljak, S. Ferraro, E. Castorina, and especially Alex Kim for helpful comments on the draft.

-
- [1] P. J. E. Peebles, *Principles of Physical Cosmology* by P.J.E. Peebles. Princeton University Press, 1993. ISBN: 978-0-691-01933-8 (1993).
 - [2] J. R. Bond, A. H. Jaffe, and L. Knox, “Estimating the power spectrum of the cosmic microwave background,” *Phys. Rev. D* **57**, 2117–2137 (1998).

- [3] Ue-Li Pen, “Fast power spectrum estimation,” *Mon. Not. Roy. Astron. Soc.* **346**, 619–626 (2003), arXiv:astro-ph/0304513 [astro-ph].
- [4] N. Padmanabhan, U. Seljak, and U. L. Pen, “Mining weak lensing surveys.” *New Astronomy* **8**, 581–603 (2003).
- [5] Benjamin D. Wandelt, David L. Larson, and Arun Lakshminarayanan, “Global, exact cosmic microwave background data analysis using Gibbs sampling,” *Phys. Rev. D* **70**, 083511 (2004), arXiv:astro-ph/0310080 [astro-ph].
- [6] Kendrick M. Smith, Oliver Zahn, and Olivier Doré, “Detection of gravitational lensing in the cosmic microwave background,” *Phys. Rev. D* **76**, 043510 (2007), arXiv:0705.3980 [astro-ph].
- [7] F. S. Kitaura and T. A. Enßlin, “Bayesian reconstruction of the cosmological large-scale structure: methodology, inverse algorithms and numerical optimization,” *Mon. Not. Roy. Astron. Soc.* **389**, 497–544 (2008), arXiv:0705.0429 [astro-ph].
- [8] Uroš Seljak, Grigor Aslanyan, Yu Feng, and Chirag Modi, “Towards optimal extraction of cosmological information from nonlinear data,” *Journal of Cosmology and Astro-Particle Physics* **2017**, 009 (2017).
- [9] A. Font-Ribera, P. McDonald, and A. Slosar, “How to estimate the 3D power spectrum of the Lyman- α forest,” *JCAP* **1**, 003 (2018), arXiv:1710.11036.
- [10] N. Aghanim, M. Arnaud, M. Ashdown, J. Aumont, C. Baccigalupi, A. J. Banday, R. B. Barreiro, J. G. Bartlett, N. Bartolo, E. Battaner, K. Benabed, A. Benoît, A. Benoit-Lévy, J. P. Bernard, M. Bersanelli, P. Bielewicz, J. J. Bock, A. Bonaldi, L. Bonavera, J. R. Bond, J. Borrill, F. R. Bouchet, F. Boulanger, M. Bucher, C. Burigana, R. C. Butler, E. Calabrese, J. F. Cardoso, A. Catalano, A. Challinor, H. C. Chiang, P. R. Christensen, D. L. Clements, L. P. L. Colombo, C. Combet, A. Coulais, B. P. Crill, A. Curto, F. Cuttaia, L. Danese, R. D. Davies, R. J. Davis, P. de Bernardis, A. de Rosa, G. de Zotti, J. Delabrouille, F. X. Désert, E. Di Valentino, C. Dickinson, J. M. Diego, K. Dolag, H. Dole, S. Donzelli, O. Doré, M. Douspis, A. Ducout, J. Dunkley, X. Dupac, G. Efstathiou, F. Elsner, T. A. Enßlin, H. K. Eriksen, J. Fergusson, F. Finelli, O. Forni, M. Frailis, A. A. Fraisse, E. Franceschi, A. Frejsel, S. Galeotta, S. Galli, K. Ganga, C. Gauthier, M. Gerbino, M. Giard, E. Gjerløw, J. González-Nuevo, K. M. Górski, S. Gratton, A. Gregorio, A. Gruppuso, J. E. Gudmundsson, J. Hamann, F. K. Hansen, D. L. Harrison, G. Helou, S. Henrot-Versillé, C. Hernández-Monteagudo, D. Herranz, S. R. Hildebrandt, E. Hivon, W. A. Holmes, A. Hornstrup, K. M. Huffenberger, G. Hurier, A. H. Jaffe, W. C. Jones, M. Juvela, E. Keihänen, R. Keskitalo, K. Kiiveri, J. Knoche, L. Knox, M. Kunz, H. Kurki-Suonio, G. Lagache, A. Lähteenmäki, J. M. Lamarre, A. Lasenby, M. Lattanzi, C. R. Lawrence, M. Le Jeune, R. Leonardi, J. Lesgourgues, F. Levrier, A. Lewis, M. Liguori, P. B. Lilje, M. Lilley, M. Linden-Vørnle, V. Lindholm, M. López-Cañiego, J. F. Macías-Pérez, B. Maffei, G. Maggio, D. Maino, N. Mandolesi, A. Mangilli, M. Maris, P. G. Martin, E. Martínez-González, S. Masi, S. Matarrese, P. R. Meinhold, A. Melchiorri, M. Migliaccio, M. Millea, S. Mitra, M. A. Miville-Deschênes, A. Moneti, L. Montier, G. Morgante, D. Mortlock, S. Mottet, D. Munshi, J. A. Murphy, A. Narimani, P. Naselsky, F. Nati, P. Natoli, F. Novello, D. Novikov, I. Novikov, C. A. Oxborrow, F. Paci, L. Pagano, F. Pajot, D. Paoletti, B. Partridge, F. Pasian, G. Patanchon, T. J. Pearson, O. Perdereau, L. Perotto, V. Pettorino, F. Piacentini, M. Piat, E. Pierpaoli, D. Pietrobon, S. Plaszczynski, E. Pointecouteau, G. Polenta, N. Ponthieu, G. W. Pratt, S. Prunet, J. L. Puget, J. P. Rachen, M. Reinecke, M. Remazeilles, C. Renault, A. Renzi, I. Ristorcelli, G. Rocha, M. Rossetti, G. Roudier, B. Rouillé d’Orfeuil, J. A. Rubiño-Martín, B. Rusholme, L. Salvati, M. Sandri, D. Santos, M. Savelainen, G. Savini, D. Scott, P. Serra, L. D. Spencer, M. Spinelli, V. Stolyarov, R. Stompor, R. Sunyaev, D. Sutton, A. S. Suur-Uski, J. F. Sygnet, J. A. Tauber, L. Terenzi, L. Toffolatti, M. Tomasi, M. Tristram, T. Trombetti, M. Tucci, J. Tuovinen, G. Umata, L. Valenziano, J. Valiviita, F. Van Tent, P. Vielva, F. Villa, L. A. Wade, B. D. Wandelt, I. K. Wehus, D. Yvon, A. Zacchei, and A. Zonca, “Planck 2015 results. XI. CMB power spectra, likelihoods, and robustness of parameters,” *Astron. Astrophys.* **594**, A11 (2016).
- [11] F. Beutler, H.-J. Seo, A. J. Ross, P. McDonald, S. Saito, A. S. Bolton, J. R. Brownstein, C.-H. Chuang, A. J. Cuesta, D. J. Eisenstein, A. Font-Ribera, J. N. Grieb, N. Hand, F.-S. Kitaura, C. Modi, R. C. Nichol, W. J. Percival, F. Prada, S. Rodríguez-Torres, N. A. Roe, N. P. Ross, S. Salazar-Albornoz, A. G. Sánchez, D. P. Schneider, A. Slosar, J. Tinker, R. Tojeiro, M. Vargas-Magaña, and J. A. Vazquez, “The clustering of galaxies in the completed SDSS-III Baryon Oscillation Spectroscopic Survey: baryon acoustic oscillations in the Fourier space,” *Mon. Not. Roy. Astron. Soc.* **464**, 3409–3430 (2017), arXiv:1607.03149.
- [12] K. G. Wilson and J. Kogut, “The renormalization group and the ϵ expansion,” *Phys. Rept.* **12**, 75–199 (1974).
- [13] T. Banks, *Modern Quantum Field Theory, by Tom Banks, Cambridge, UK: Cambridge University Press, 2008* (2008).
- [14] One may wonder how generic is this behavior that differences in χ^2 between models have much better accuracy than the absolute value of χ^2 . In practice it does not really matter because we will always test numerical convergence of the quantities of interest (generally differences between models) on a case-by-case basis, determining whether they converge without regard to whether they converge faster than absolute χ^2 or not, but maybe it is an interesting academic question. I think this behavior probably is generic, because of the usual relation between absolute χ^2 , degrees of freedom (dof), and $\Delta\chi^2$ between models. Absolute χ^2 for any data set has big random fluctuations around a big number dof, much more for a large data set than the level of $\Delta\chi^2$ we would call significant as a difference between models (e.g., we would say a model is ruled out decisively if it fits $\Delta\chi^2 = 25$ worse than another one, even if both of them would look like they had a “fine χ^2 ”, i.e., $\chi^2 \sim \text{dof}$, if compared to that data in isolation). The reason these things work this way is that most of the contribution to absolute χ^2 is effectively orthogonal to any smooth model, so it is irrelevant to comparison between models. I think we can generically expect this likelihood calculation to behave similarly. Again though, we can/must always do convergence tests to verify that the likelihood differences that we care about are not sensitive to numerical parameters, so we won’t ever be fooled by imperfection in this intuition.
- [15] P. McDonald, “Primordial non-Gaussianity: Large-scale structure signature in the perturbative bias model,” *Phys. Rev. D* **78**, 123519+ (2008), arXiv:0806.1061.
- [16] N. Bartolo, J. P. Beltrán Almeida, S. Matarrese, M. Pietroni, and A. Riotto, “Signatures of primordial non-Gaussianities in the matter power-spectrum and bispectrum: the time-RG approach,” *JCAP* **3**, 011 (2010), arXiv:0912.4276 [astro-ph.CO].

- [17] T. Giannantonio and C. Porciani, “Structure formation from non-Gaussian initial conditions: Multivariate biasing, statistics, and comparison with N-body simulations,” *Phys. Rev. D* **81**, 063530 (2010), arXiv:0911.0017 [astro-ph.CO].
- [18] J.-O. Gong and S. Yokoyama, “Scale-dependent bias from primordial non-Gaussianity with trispectrum,” *Mon. Not. Roy. Astron. Soc.* **417**, L79–L82 (2011), arXiv:1106.4404 [astro-ph.CO].
- [19] M. Alvarez, T. Baldauf, J. R. Bond, N. Dalal, R. de Putter, O. Doré, D. Green, C. Hirata, Z. Huang, D. Huterer, D. Jeong, M. C. Johnson, E. Krause, M. Loverde, J. Meyers, P. D. Meerburg, L. Senatore, S. Shandera, E. Silverstein, A. Slosar, K. Smith, M. Zaldarriaga, V. Assassi, J. Braden, A. Hajian, T. Kobayashi, G. Stein, and A. van Engelen, “Testing Inflation with Large Scale Structure: Connecting Hopes with Reality,” ArXiv e-prints (2014), arXiv:1412.4671.
- [20] Azadeh Moradinezhad Dizgah and Cora Dvorkin, “Scale-dependent galaxy bias from massive particles with spin during inflation,” *Journal of Cosmology and Astro-Particle Physics* **2018**, 010 (2018).
- [21] P. Valageas, “Dynamics of gravitational clustering. I. Building perturbative expansions,” *Astron. Astrophys.* **379**, 8–20 (2001).
- [22] P. Valageas, “Dynamics of gravitational clustering. II. Steepest-descent method for the quasi-linear regime,” *Astron. Astrophys.* **382**, 412–430 (2002), astro-ph/0107126.
- [23] P. Valageas, “A new approach to gravitational clustering: A path-integral formalism and large-N expansions,” *Astron. Astrophys.* **421**, 23–40 (2004), astro-ph/0307008.
- [24] P. Valageas, “Large-N expansions applied to gravitational clustering,” *Astron. Astrophys.* **465**, 725–747 (2007), astro-ph/0611849.
- [25] P. McDonald, “Dark matter clustering: A simple renormalization group approach,” *Phys. Rev. D* **75**, 043514+ (2007).
- [26] P. McDonald, “How to generate a significant effective temperature for cold dark matter, from first principles,” *JCAP* **4**, 32+ (2011), arXiv:0910.1002 [astro-ph.CO].
- [27] U. Seljak and P. McDonald, “Distribution function approach to redshift space distortions,” *JCAP* **11**, 039 (2011), arXiv:1109.1888 [astro-ph.CO].
- [28] D. Baumann, A. Nicolis, L. Senatore, and M. Zaldarriaga, “Cosmological non-linearities as an effective fluid,” *JCAP* **7**, 051 (2012), arXiv:1004.2488 [astro-ph.CO].
- [29] J. J. M. Carrasco, M. P. Hertzberg, and L. Senatore, “The effective field theory of cosmological large scale structures,” *Journal of High Energy Physics* **9**, 82 (2012), arXiv:1206.2926 [astro-ph.CO].
- [30] M. P. Hertzberg, “Effective field theory of dark matter and structure formation: Semianalytical results,” *Phys. Rev. D* **89**, 043521 (2014).
- [31] P. McDonald, “What the “simple renormalization group” approach to dark matter clustering really was,” ArXiv e-prints (2014), arXiv:1403.7235.
- [32] D. Blas, S. Floerchinger, M. Garny, N. Tetradis, and U. A. Wiedemann, “Large scale structure from viscous dark matter,” *JCAP* **11**, 049 (2015), arXiv:1507.06665.
- [33] G. Rigopoulos, “The adhesion model as a field theory for cosmological clustering,” *JCAP* **1**, 014 (2015), arXiv:1404.7283.
- [34] F. F  hrer and G. Rigopoulos, “Renormalizing a viscous fluid model for large scale structure formation,” *JCAP* **2**, 032 (2016), arXiv:1509.03073.
- [35] M. Bartelmann, F. Fabis, D. Berg, E. Kozlikin, R. Lilow, and C. Viermann, “A microscopic, non-equilibrium, statistical field theory for cosmic structure formation,” *New Journal of Physics* **18**, 043020 (2016).
- [36] M. Schmittfull, Z. Vlah, and P. McDonald, “Fast large scale structure perturbation theory using one-dimensional fast Fourier transforms,” *Phys. Rev. D* **93**, 103528 (2016), arXiv:1603.04405.
- [37] Y. Feng, M.-Y. Chu, U. Seljak, and P. McDonald, “FASTPM: a new scheme for fast simulations of dark matter and haloes,” *Mon. Not. Roy. Astron. Soc.* **463**, 2273–2286 (2016), arXiv:1603.00476.
- [38] Patrick McDonald and Zvonimir Vlah, “Large-scale structure perturbation theory without losing stream crossing,” *Phys. Rev. D* **97**, 023508 (2018).
- [39] Goran Jelic-Cizmek, Francesca Lepori, Julian Adamek, and Ruth Durrer, “The generation of vorticity in cosmological N-body simulations,” *Journal of Cosmology and Astro-Particle Physics* **2018**, 006 (2018).
- [40] Massimo Pietroni, “Structure formation beyond shell-crossing: nonperturbative expansions and late-time attractors,” *Journal of Cosmology and Astro-Particle Physics* **2018**, 028 (2018).
- [41] Enrico Pajer and Drian van der Woude, “Divergence of perturbation theory in large scale structures,” *Journal of Cosmology and Astro-Particle Physics* **2018**, 039 (2018).
- [42] Felix Fabis, Elena Kozlikin, Robert Lilow, and Matthias Bartelmann, “Kinetic field theory: exact free evolution of Gaussian phase-space correlations,” *Journal of Statistical Mechanics: Theory and Experiment* **4**, 043214 (2018).
- [43] Michael Kopp, Kyriakos Vattis, and Constantinos Skordis, “Solving the Vlasov equation in two spatial dimensions with the Schr  dinger method,” *Phys. Rev. D* **96**, 123532 (2017).
- [44] Atsushi Taruya, Takahiro Nishimichi, and Donghui Jeong, “GridSPT: Grid-based calculation for perturbation theory of large-scale structure,” ArXiv e-prints, arXiv:1807.04215 (2018), arXiv:1807.04215 [astro-ph.CO].
- [45] E. E. Jenkins, A. V. Manohar, W. J. Waalewijn, and A. P. S. Yadav, “Higher-order gravitational lensing reconstruction using Feynman diagrams,” *JCAP* **9**, 024 (2014), arXiv:1403.4607.
- [46] M. S. Madhavacheril, P. McDonald, N. Sehgal, and A. Slosar, “Building unbiased estimators from non-Gaussian likelihoods with application to shear estimation,” *JCAP* **1**, 022 (2015), arXiv:1407.1906.
- [47] E. Di Dio, “Lensing smoothing of BAO wiggles,” *JCAP* **3**, 016 (2017), arXiv:1609.09044.
- [48] P. McDonald, “Clustering of dark matter tracers: Renormalizing the bias parameters,” *Phys. Rev. D* **74**, 103512+ (2006).
- [49] R. E. Smith, R. Scoccimarro, and R. K. Sheth, “Scale dependence of halo and galaxy bias: Effects in real space,” *Phys. Rev. D* **75**, 063512+ (2007), arXiv:astro-ph/0609547.

- [50] T. Matsubara, “Nonlinear perturbation theory with halo bias and redshift-space distortions via the Lagrangian picture,” *Phys. Rev. D* **78**, 083519+ (2008).
- [51] P. McDonald and A. Roy, “Clustering of dark matter tracers: generalizing bias for the coming era of precision LSS,” *JCAP* **8**, 020 (2009), arXiv:0902.0991 [astro-ph.CO].
- [52] D. Jeong and E. Komatsu, “Perturbation Theory Reloaded. II. Nonlinear Bias, Baryon Acoustic Oscillations, and Millennium Simulation in Real Space,” *Astrophys. J.* **691**, 569–595 (2009), arXiv:0805.2632.
- [53] T. Matsubara, “Nonlinear perturbation theory integrated with nonlocal bias, redshift-space distortions, and primordial non-Gaussianity,” *Phys. Rev. D* **83**, 083518 (2011), arXiv:1102.4619 [astro-ph.CO].
- [54] A. Elia, S. Kulkarni, C. Porciani, M. Pietroni, and S. Matarrese, “Modelling the clustering of dark matter haloes in resummed perturbation theories,” *Mon. Not. Roy. Astron. Soc.* **416**, 1703–1716 (2011), arXiv:1012.4833 [astro-ph.CO].
- [55] K. C. Chan and R. Scoccimarro, “Halo sampling, local bias, and loop corrections,” *Phys. Rev. D* **86**, 103519 (2012), arXiv:1204.5770 [astro-ph.CO].
- [56] T. Baldauf, U. Seljak, V. Desjacques, and P. McDonald, “Evidence for quadratic tidal tensor bias from the halo bispectrum,” *Phys. Rev. D* **86**, 083540 (2012), arXiv:1201.4827 [astro-ph.CO].
- [57] T. Nishimichi and A. Oka, “Simulating the anisotropic clustering of luminous red galaxies with subhaloes: a direct confrontation with observation and cosmological implications,” *Mon. Not. Roy. Astron. Soc.* **444**, 1400–1418 (2014), arXiv:1310.2672.
- [58] S. Saito, T. Baldauf, Z. Vlah, U. Seljak, T. Okumura, and P. McDonald, “Understanding higher-order nonlocal halo bias at large scales by combining the power spectrum with the bispectrum,” *Phys. Rev. D* **90**, 123522 (2014), arXiv:1405.1447.
- [59] V. Desjacques, D. Jeong, and F. Schmidt, “Large-Scale Galaxy Bias,” *ArXiv e-prints* (2016), arXiv:1611.09787.
- [60] N. Hand, U. Seljak, F. Beutler, and Z. Vlah, “Extending the modeling of the anisotropic galaxy power spectrum to $k = 0.4 \, h\text{Mpc}^{-1}$,” *ArXiv e-prints* (2017), arXiv:1706.02362.
- [61] C.-H. Chuang, F.-S. Kitaura, Y. Liang, A. Font-Ribera, C. Zhao, P. McDonald, and C. Tao, “Linear redshift space distortions for cosmic voids based on galaxies in redshift space,” *Phys. Rev. D* **95**, 063528 (2017), arXiv:1605.05352.
- [62] M. Ata, F.-S. Kitaura, C.-H. Chuang, S. Rodríguez-Torres, R. E. Angulo, S. Ferraro, H. Gil-Marín, P. McDonald, C. Hernández Monteagudo, V. Müller, G. Yepes, M. Autefage, F. Baumgarten, F. Beutler, J. R. Brownstein, A. Burden, D. J. Eisenstein, H. Guo, S. Ho, C. McBride, M. Neyrinck, M. D. Olmstead, N. Padmanabhan, W. J. Percival, F. Prada, G. Rossi, A. G. Sánchez, D. Schlegel, D. P. Schneider, H.-J. Seo, A. Strebljanska, J. Tinker, R. Tojeiro, and M. Vargas-Magana, “The clustering of galaxies in the completed SDSS-III Baryon Oscillation Spectroscopic Survey: cosmic flows and cosmic web from luminous red galaxies,” *Mon. Not. Roy. Astron. Soc.* **467**, 3993–4014 (2017), arXiv:1605.09745.
- [63] T. A. Enßlin, M. Frommert, and F. S. Kitaura, “Information field theory for cosmological perturbation reconstruction and nonlinear signal analysis,” *Phys. Rev. D* **80**, 105005 (2009), arXiv:0806.3474.
- [64] E. Di Dio, H. Perrier, R. Durrer, G. Marozzi, A. Moradinezhad Dizgah, J. Noreña, and A. Riotto, “Non-Gaussianities due to relativistic corrections to the observed galaxy bispectrum,” *Journal of Cosmology and Astro-Particle Physics* **2017**, 006 (2017).
- [65] A. Font-Ribera, P. McDonald, N. Mostek, B. A. Reid, H.-J. Seo, and A. Slosar, “DESI and other Dark Energy experiments in the era of neutrino mass measurements,” *JCAP* **5**, 023 (2014), arXiv:1308.4164.
- [66] DESI Collaboration, A. Aghamousa, J. Aguilar, S. Ahlen, S. Alam, L. E. Allen, C. Allende Prieto, J. Annis, S. Bailey, C. Balland, and et al., “The DESI Experiment Part I: Science, Targeting, and Survey Design,” *ArXiv e-prints* (2016), arXiv:1611.00036 [astro-ph.IM].
- [67] S. Matarrese and M. Pietroni, “Resumming cosmic perturbations,” *Journal of Cosmology and Astro-Particle Physics* **6**, 26+ (2007), arXiv:astro-ph/0703563.
- [68] K. Izumi and J. Soda, “Renormalized Newtonian cosmic evolution with primordial non-Gaussianity,” *Phys. Rev. D* **76**, 083517+ (2007), arXiv:0706.1604.
- [69] O. J. Rosten, “A comment on the path integral approach to cosmological perturbation theory,” *JCAP* **1**, 029 (2008), arXiv:0711.0867.
- [70] S. Matarrese and M. Pietroni, “Baryonic Acoustic Oscillations via the Renormalization Group,” *Modern Physics Letters A* **23**, 25–32 (2008), arXiv:astro-ph/0702653.
- [71] S. Floerchinger, M. Garny, N. Tetradis, and U. A. Wiedemann, “Renormalization-group flow of the effective action of cosmological large-scale structures,” *JCAP* **1**, 048 (2017), arXiv:1607.03453.
- [72] Torsten A. Enßlin, “Information theory for fields,” *ArXiv e-prints*, arXiv:1804.03350 (2018), arXiv:1804.03350 [astro-ph.CO].